**The EMBO Journal** Vol. 21 No. 4 pp. 801–814, 2002

# Testing water-mediated DNA recognition by the Hin recombinase

**Thang Kien Chiu[1,2], Catherine Sohn[3], Richard E.Dickerson[1,4,5] and Reid C.Johnson[3,4,5]**

[1]Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, CA 90095, [3]Department of Biological Chemistry, UCLA School of Medicine, Los Angeles, CA 90095-1737 and [4]Molecular Biology Institute, University of California at Los Angeles, Los Angeles, CA 90095-1570, USA

[2]Present address: Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20892, USA

[5]Corresponding authors: Reid C.Johnson, Department of Biological Chemistry, UCLA School of Medicine, Los Angeles, CA 90095-1737, USA or Richard E.Dickerson, Molecular Biology Institute, University of California, Los Angeles, CA 90095-1570, USA
e-mail: rcjohnson@mednet.ucla.edu or red@mbi.ucla.edu

**The Hin recombinase specifically recognizes its DNA-binding site by means of both major and minor groove interactions. A previous X-ray structure, together with new structures of the Hin DNA-binding domain bound to a recombination half-site that were solved as part of the present study, have revealed that two ordered water molecules are present within the major groove interface. In this report, we test the importance of these waters directly by X-ray crystal structure analysis of complexes with four mutant DNA sequences. These structures, combined with their Hin-binding properties, provide strong support for the critical importance of one of the intermediate waters. A lesser but demonstrable role is ascribed to the second water molecule. The mutant structures also illustrate the prominent roles of thymine methyls both in stabilizing intermediate waters and in interfering with water or amino acid side chain interactions with DNA.**

*Keywords*: DNA invertase/helix–turn–helix motif/protein–DNA recognition/water/X-ray crystallography

## Introduction

The reading of DNA sequences by recognition proteins involves two primary factors: (i) the pattern of H-bond donors and acceptors and van der Waals contacts along the floor of the major or minor groove (Seeman *et al*., 1976; von Hippel and Berg, 1986); and (ii) local deformations of helix structure that are intrinsic to the base sequence or induced by protein binding (Dickerson and Chiu, 1997; Olson *et al*., 1998). Recognition via H-bonding involves atoms of the protein and the DNA, with water molecules sometimes playing an intermediary role (Schwabe, 1997; Woda *et al*., 1998).

Recent surveys of high-resolution protein–DNA crystal structures have noted that solvent molecules are commonly present within protein–DNA interfaces (Nadassy *et al*., 1999, 2001; Luscombe *et al*., 2001). Indeed, water-mediated interactions can be as common as direct H-bonds or salt links. Of these, >70% mediate interactions between the DNA backbone and protein. The remainder are located between DNA base and protein residues, but of these only 20% (or <1% of overall protein–DNA interactions) have three or more H-bonding partners (Luscombe *et al*., 2001). Hence, most of these waters are believed to be primarily functioning as space fillers.

The relative contribution of specific waters to specificity and affinity in protein–DNA recognition has seldom been addressed directly, although their importance has been inferred from structures of a number of transcription factors and enzymes in complex with DNA. One of the best studied examples of protein–DNA recognition involving water is the Trp repressor DNA complex, where water-mediated H-bonds are observed to an AG dinucleotide within the recognition sequence (Otwinowski *et al*., 1988; Lawson and Carey, 1993). Mutagenesis of the highly conserved G to an A, or of the less conserved A to a G, reduces binding affinities to 0.2 or 8.3% of their wild-type values, respectively (Joachimiak *et al*., 1994). The deleterious effect of the G-to-A mutation is reversed when it is combined with the A-to-G mutation, with the binding affinity of the double mutant being 5.7% of wild type. Hydrogen bond donor and acceptor atoms of the DNA bases are interchanged in the double mutant, and it has been suggested that two or three water molecules could still recognize the changed configuration (Joachimiak *et al*., 1994). However, the presence and roles of these waters have not yet been tested by a structural analysis of mutant Trp repressor–operator complexes.

Other examples pointing to the importance of water include DNA complexes of related homeodomain proteins where particular waters have been observed at common locations (Gehring *et al*., 1994; Hirsch and Aggarwal, 1995; Li *et al*., 1995; Wilson *et al*., 1995; Wolberger, 1996; Fraenkel and Pabo, 1998; Fraenkel *et al*., 1998). On the other hand, considerable flexibility in the disposition of particular solvent molecules within the protein–DNA interface can exist. A conserved glutamine within the homeodomain recognition helix adopts multiple conformations that may or may not involve indirect base contacts via water. This glutamine is believed to be important for sequence-specific recognition, but it has been difficult to define its structural role precisely. Mutations at this glutamine in Engrailed have varied affects on binding. In the case of an alanine substitution for this residue in Engrailed, several ordered waters essentially replace the glutamine side chain atoms with little consequence on binding (Grant *et al*., 2000). On the other hand, a lysine at this position changes specificity in favor of an alternate sequence where primarily direct base contacts are formed (Tucker-Kellogg *et al*., 1997). The restriction

endonuclease *Eco*RV has also been shown to adopt different patterns of direct and water-mediated interactions with bases immediately outside the recognition sequence, depending on the identity of the flanking nucleotides (Horton and Perona, 1998). These alternate contacts may contribute to the variations in binding constants measured at different cleavage sites.

We have previously reported that the DNA complex of the Hin recombinase DNA-binding domain (Hin-DBD) contains two ordered water molecules, which were proposed to play an important role in selective recognition within the major groove (Feng *et al*., 1994). Hin catalyzes a site-specific DNA inversion reaction within the chromosome of *Salmonella typhimurium* (Silverman *et al*., 1981; Johnson, 2002). Inversion of the 1 kb DNA segment between the *hixL* and *hixR* recombination sites switches the orientation of a promoter that controls the alternate synthesis of two different flagellins. Hin is a member of an extended family of serine recombinases, which includes other closely related DNA invertases as well as the more distantly related DNA resolvases (Grindley, 1993; Johnson, 2002).

In the present study, we directly test the importance of the two water molecules within the major groove interface in binding by relating the X-ray structures of complexes containing the Hin-DBD bound to four mutant sites with their *in vitro* binding properties. We find that one of the waters is an essential determinant for sequence-specific binding by Hin. The presence of this water is strictly correlated with stable complex formation, although the atomic interactions vary with the identity of the base. The second water appears to play a more auxiliary role. This study represents the first systematic analysis accounting for the structural contributions of water molecules in sequence-specific protein–DNA recognition. The mutant structures also illuminate the impact of thymine methyls on the major groove surface on stabilizing or disrupting intermediary waters or direct amino acid contacts.

## Results

### *Crystal structures of Hin complexes to wild-type and four mutant binding sites*

A synthetically derived peptide consisting of the C-terminal 52 amino acids was crystallized with duplex DNAs of 13 bp plus a single 5′ overhang at each end representing the right half of the mutant *hixL* sequences (Figure 1A and B). During the course of this work, the structure of the Hin-DBD bound to the wild-type *hixL* half-site, designated Nat-1, was redetermined using newly obtained synchroton data. Crystals listed as 'Form 1' in Table I were isormorphous with the previously determined wild-type structure Nat-0 (Feng *et al*., 1994), but were obtained from different crystallization conditions. Phases were determined by multiple isomorphous replacement plus anomalous scattering, using DNA sequences derivatized at bases T4, T5, C18 or T7+T19. The Br18 and I5 derivatives were chosen for refinement as the 'native' Form 1 structures because of the higher quality and resolution of their data sets, and Br18 is the designated reference wild-type structure in the following discussion. The new phases were extended to four mutant structures: G9T, A10G, T11G and T11C. Among the Form 1 structures of Table I, root-mean-square differences (r.m.s.ds) from the reference structure Br18 for residues that they have in common range from 0.4 to 0.8 Å (Table II), only slightly higher than the estimated coordinate error of ~0.4 Å for each structure.

The wild-type Hin–DNA complex was also solved in a different packing mode ('Form 2') that was produced under different crystallization conditions, both with (I4-2) and without (Nat-2) iodination of base T4. The r.m.s.d. between common features in the Form 1 (Br18) and Form 2 (I4-2) models (0.8 Å) indicates that the two structures are not significantly different. Whereas I4-2 was useful in confirming aspects of the structure of the complex, it is not considered below in the discussion of major groove interactions mediated by water molecules because of its lower resolution.

The molecular models for Nat-0 and Br18 differ in two respects, neither of which relates to the issue of water-mediated sequence recognition: (i) details of stacking of DNA helices within the crystal lattice have been reinterpreted (Chiu, 2001); and (ii) residues 182–185 following helix 3 are now traced as a coiled conformation extending away from the DNA, similar to the tail of γδ resolvase (Yang and Steitz, 1995), and as opposed to the earlier model where the C-terminal residues 186–190 resided within the DNA minor groove. The C-terminal 3–7 residues in the current models are disordered and not observed. The present structure of the C-terminus is observed in both Form 1 and Form 2. This assignment is consistent with observations that residues 186–190 are not present in all DNA invertases, and that residues following 184 in Hin can be deleted without affecting DNA binding (R.Johnson, unpublished data; K.Hughes, personal communication).

**Fig. 1.** Structure of the Hin-DBD–DNA complex. (**A**) Sequence of the duplex oligonucleotide used for crystallography representing the wild-type *hixL* inside half-site. Asterisks denote positions where methylation by dimethyl sulfate inhibits Hin binding (Glasgow *et al*., 1989). (**B**) Amino acid sequence of the C-terminal 52 amino acid Hin DNA-binding domain used for crystal growth. Helices are denoted by rectangles. (**C**) Stereo diagram of the Hin-DBD bound to the *hix* half-site as represented by structure Br18. The N-terminus of the three-helix bundle (amino acid residues 139–143) slips into the minor groove, while the C-terminal chain (182–185) extends away from the DNA. Helices are rendered as cylinders. Main or side chain atoms that H-bond with the DNA backbone are green, and side chains that H-bond to base atoms or water (gray spheres) are rendered as red ball and sticks. A:T/T:A base pairs are colored in alternating shades of blue and C:G/G:C base pairs are colored magenta. Hydrogen bonding is depicted with dashed lines. (**D**) Schematic ladder diagram of protein–DNA interactions. The major groove is at the upper left and the minor groove is at the lower right. Blue ovals represent thymine methyl groups and gray spheres are waters. Filled black circles represent sugars, and phosphate oxygens are located along the lines that connect the sugars. Solid arrows denote H-bonds, pointing from donor to acceptor. Dashed lines are close van der Waals contacts of ⩽3.5 Å. Amino acid residues making base contacts are red, and those making DNA backbone contacts are black. (**E**) Stereo view of minor groove interactions in the wild-type Br18 structure. Gly139 (yellow) and Arg140 (magenta) are shown with their van der Waals surfaces, along with the main chain segment extending to Ala143. Oxygen and nitrogen atoms of residues 139–140 are blue and red, respectively. Black dotted lines are H-bonds and red dotted lines are close van der Waals contacts.

### *Overall view of the Hin–DNA complex*

The wild-type (Br18) structure and an unrolled ladder diagram of all direct and indirect contacts between the protein and DNA are shown in Figure 1C and D. The DNA-binding domain of Hin is arranged in a compact helix–turn–helix motif consisting of three short α-helices

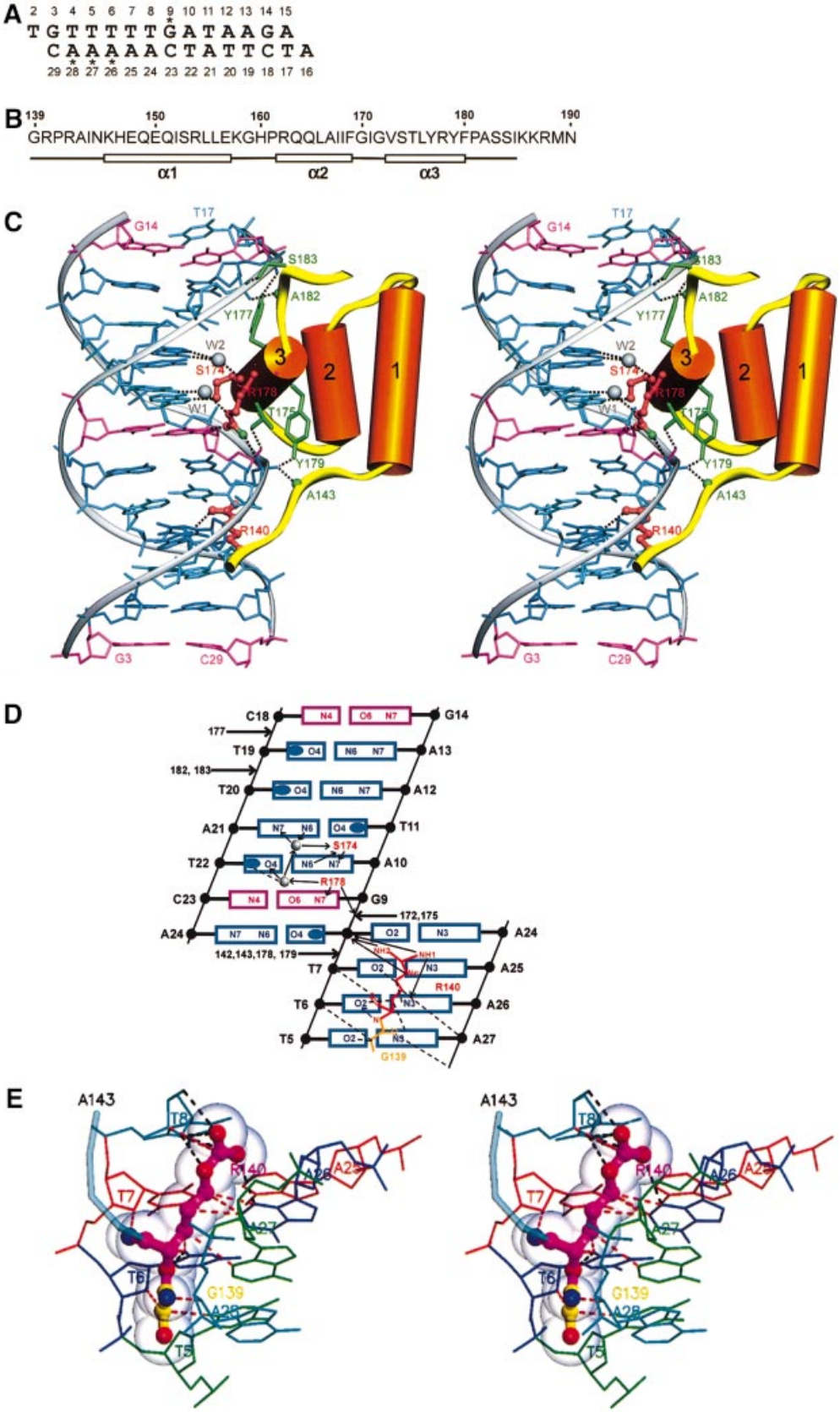


**Table I.** Data collection and heavy atom phasing statistics

| Crystal form | Form 1 | | | | | | | | | | | Form 2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DNA sequence | WT sequence and its heavy atom derivatives | | | | | | Mutants | | | | | WT and derivatives | |
| Structure ID | Nat-1 | I7/19 | I4 | I4 | I5 | I5 | Br18 | G9T | A10G | T11G | T11C | Nat-2 | I4-2 |
| Data collection (space group = $C222_1$) | | | | | | | | | | | | | |
| Wavelength (Å) | 0.908 | 1.100 | 1.543 | 0.920 | 1.543 | 0.980 | 0.917 | 1.100 | 1.100 | 1.100 | 0.908 | 1.543 | 1.543 |
| Resolution (Å) | 2.20 | 2.60 | 2.80 | 2.40 | 2.70 | 2.30 | 2.40 | 2.86 | 2.24 | 2.83 | 2.28 | 2.90 | 2.75 |
| No. of reflections | | | | | | | | | | | | | |
| unique | 6583 | 4858 | 3822 | 5470 | 7959 | 6695 | 5435 | 3502 | 6235 | 3774 | 7102 | 3168 | 3765 |
| redundancy | 12 | 8 | 10 | 21 | 10 | 43 | 26 | 12 | 15 | 13 | 17 | 10 | 15 |
| Completeness (%) | | | | | | | | | | | | | |
| overall | 77.4 | 95.2 | 88.5 | 84.5 | 94.4 | 89.6 | 85.4 | 89.6 | 81.3 | 91.2 | 93.2 | 90.7 | 96.5 |
| last shell | 42.0 | 82.0 | 56.9 | 40.7 | 73.1 | 55.3 | 43.8 | 80.8 | 50.8 | 76.4 | 72.1 | 71.8 | 91.6 |
| $R_{sym}$ (%) | | | | | | | | | | | | | |
| overall | 8.4 | 6.6 | 5.4 | 5.7 | 5.5 | 6.7 | 7.9 | 7.6 | 6.8 | 7.7 | 8.9 | 8.6 | 8.1 |
| last shell | 22.9 | 28.3 | 23.5 | 27.4 | 27.0 | 19.0 | 27.0 | 18.8 | 21.5 | 21.8 | 25.2 | 28.5 | 22.2 |
| $I/\sigma$ | | | | | | | | | | | | | |
| overall | 9.1 | 13.9 | 20.7 | 23.0 | 21.4 | 21.1 | 16.6 | 15.3 | 19.7 | 16.9 | 15.7 | 15.1 | 19.6 |
| last shell | 1.5 | 3.4 | 3.3 | 2.9 | 3.1 | 6.3 | 2.6 | 3.0 | 2.1 | 4.1 | 2.5 | 2.6 | 4.9 |
| Phasing statistics | | | | | | | | | | | | | |
| $R_{iso}$ (%) | | 20.5 | 16.9 | 15.8 | 20.3 | 18.6 | 13.7 | | | | | | 20.3 |
| $R_{ano}$ (%) | | 6.3 | 6.3 | 4.6 | 6.5 | 4.0 | 5.1 | | | | | | 5.8 |
| Phasing power | | | | | | | | | | | | | |
| centric | | 1.03 | 1.31 | 0.99 | 0.74 | 0.55 | 0.77 | | | | | | 1.14 |
| acentric | | 1.13 | 1.29 | 1.10 | 0.95 | 0.69 | 0.87 | | | | | | 1.42 |
| Figure of merit | 0.57 | | | | | | | | | | | 0.51 | |

$R_{sym} = \Sigma_i\Sigma_t|I_i - \langle I\rangle|/\Sigma_i\Sigma_t I_i$; $R_{iso} = \Sigma_h|F_{PH} - F_P|/\Sigma_h|F_P|$; $R_{ano} = \Sigma_{\pm h}|F_{PH} - \langle F_{PH}\rangle|/\Sigma_{\pm h}(F_{PH})$; phasing power = r.m.s.($F_H$)/r.m.s.(lack-of-closure error); figure of merit = $\langle P(\alpha)e^{i\phi}/P(\alpha)\rangle$, where $\alpha$ is the phase and $P(\alpha)$ is the phase probability distribution (the value given for Nat-1 is to 2.3 Å).

(Lys146–Glu157, Arg162–Phe169 and Val173–Phe180) with helix 3 inserted into the major groove. The 3-helix bundle is anchored onto the sugar–phosphate backbone of DNA residues 8, 9, 19 and 20 by contacts with nine amino acid residues. On one strand, Ala143-N (the main-chain NH) and Tyr179-OH both interact with T8-O1P, and Arg178-N$\eta$2 with T8-O2P. In addition, both Gly172-N and Thr175-O$\gamma$ are bonded to G9-O2P. On the other DNA chain, Tyr177-OH bonds with T19-O2P, while T20-O2P bonds with Ala182-N, and T20-O1P interacts with Ala182-N, Ser183-N and Ser183-O$\gamma$. These contacts form a rigid framework that holds the recognition helix 3 in the proper orientation within the major groove for reading the DNA sequence.

Specificity and affinity are enhanced by the N-terminal segment of the Hin peptide, which extends its initial two residues, Gly139 and Arg140, into the minor groove within the 'A-tract' between base pairs T5:A27 and T8:A24 (Figure 1E) (Glasgow *et al*., 1989; Sluka *et al*., 1990; Hughes *et al*., 1992). The Gly–Arg dipeptide in the N-terminus of $\gamma\delta$ resolvase is bound similarly within the minor groove (Yang and Steitz, 1995). As noted previously (Feng *et al*., 1994), the N-terminal arm and antiparallel orientation of helices 1 and 2 bear striking resemblance to the homeodomain fold. Approximately 1455–1529 Å$^2$ or 20–22% of the solvent-accessible surface of the Hin-DBD and DNA target become buried by formation of the wild-type complex, as measured by a 1.4 Å probe. The overall buried surface area was similar for each of the mutants.

### *Major groove recognition in the wild-type complexes Br18 and I5*

Recognition within the major groove involves just three base pairs, G9:C23, A10:T22 and T11:A21, which are read by two amino acid residues together with two ordered water molecules (Figure 2). The two residues are Ser174, which is invariant within DNA invertase members, and the semi-invariant Arg178 (Lys in some other DNA invertases). The waters, identified here as W1 and W2, are associated with these same two protein residues. These interface residues have a lower *B*-value than the average for the complex, supporting their importance in recognition. The H-bond pattern of G9 and A10, which are located immediately adjacent to Arg178 and Ser174, respectively, is read directly in the manner proposed earlier for sequence-specific protein–DNA recognition (Seeman *et al*., 1976; von Hippel and Berg, 1986), whereas T22 and A21, which are across the helix, are recognized through water-mediated contacts. In this respect, the waters can be thought of as extensions of the protein residues. Their presence is consistent with electrophoresis studies, which show that binding of Hin to DNA is dependent upon the osmotic pressure of the binding reaction (Robinson and Sligar, 1996).

*Ser174 and W2*. Ser174-O$\gamma$ is H-bonded directly to the acceptor A10-N7 and the donor A10-N6 (Figure 2). In the earlier model, the density for Ser174 was less well defined. As a result, Ser174 was modeled with a different rotamer, which allowed for an H-bond only to A10-N7. The conformation of Ser174 is conserved among the three

**Table II.** Refinement statistics

| Structure ID | I5 | Br18 | I4-2 | G9T | A10G | T11G | T11C |
|---|---|---|---|---|---|---|---|
| PDB code | 1IJ6 | 1IJW | 1JJ8 | 1JKQ | 1JKO | 1JKP | 1JKR |
| $R_{work}$ (%), overall/last shell | 23.8/35.6 | 22.3/42.4 | 21.3/35.3 | 25.6/39.0 | 24.3/46.6 | 24.8/41.6 | 26.9/45.4 |
| $R_{free}$ (%), overall/last shell | 28.1/38.0 | 27.9/42.1 | 27.7/51.4 | 32.8/39.5 | 30.6/43.1 | 32.8/50.2 | 32.5/44.4 |
| Average *B*-value (Å$^2$) | 52 | 40 | 45 | 53 | 64 | 54 | 60 |
| R.m.s.d. | | | | | | | |
| bonds (Å) | 0.011 | 0.009 | 0.009 | 0.009 | 0.007 | 0.009 | 0.013 |
| angles (°) | 1.9 | 1.8 | 1.3 | 1.6 | 1.3 | 1.4 | 1.7 |
| Ramachandran | | | | | | | |
| most favored | 94.9 | 92.3 | 87.8 | 74.4 | 86.8 | 87.2 | 92.1 |
| allowed | 5.1 | 7.7 | 12.2 | 25.7 | 13.2 | 12.8 | 7.9 |
| R.m.s.d. from Br18 | 0.37 | 0.00 | 0.77 | 0.55 | 0.62 | 0.52 | 0.61 |

$R_{work,\ free} = \Sigma_{hkl}|F_{obs} - F_{calc}|/\Sigma_{hkl}F_{obs}$, where 10% of the reflections were excluded during refinement for the calculation of $R_{free}$. The r.m.s.d. between Br18 and Nat-0 is 1.04 Å.

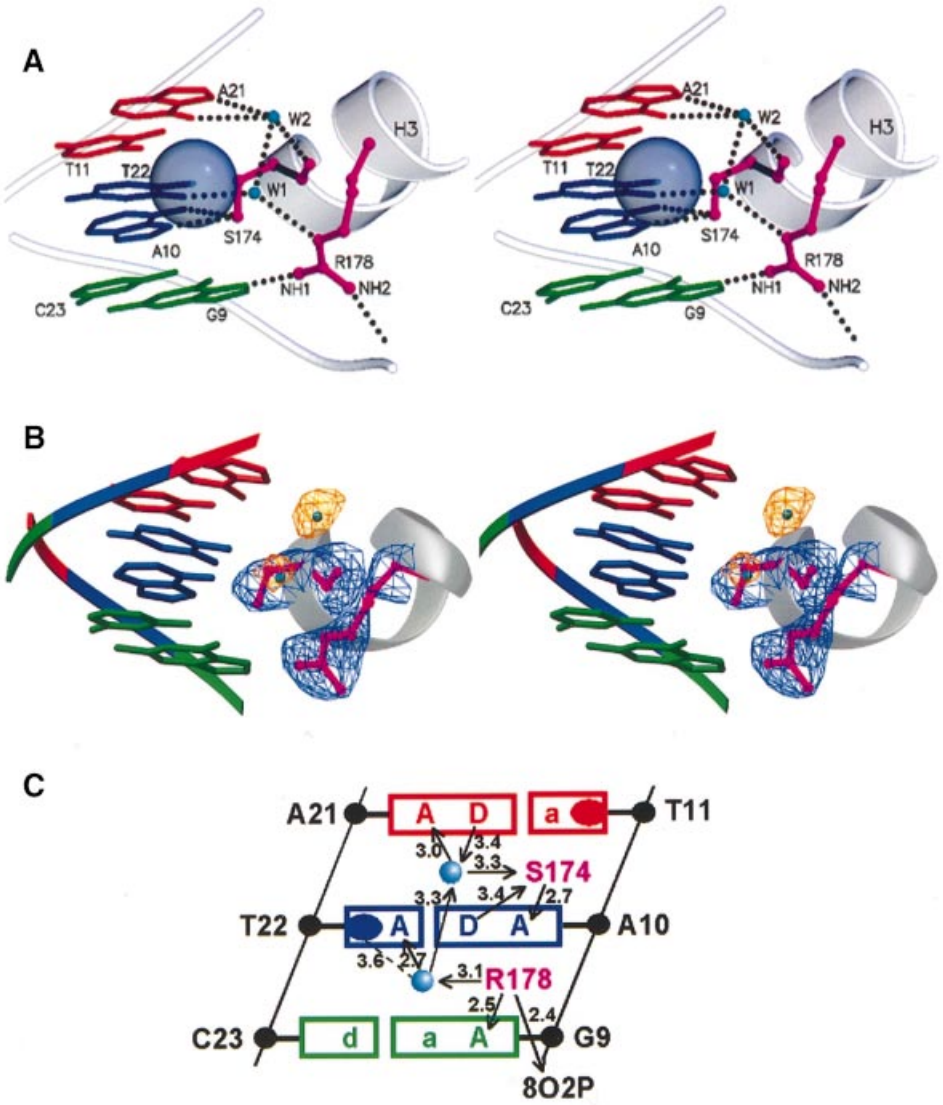



**Fig. 2.** Details of Hin–DNA interactions within the major groove of Br18. (**A**) Stereo view of major groove interactions. Ser174 and the side chain of Arg178 within helix 3 are depicted in ball-and-stick representation. The methyl of T22 is rendered in space filling to illustrate its van der Waals interaction with W1. (**B**) Simulated-anneal $F_o - F_c$ omit map of Br18 (contoured at 3.0σ), with the two interface water residues, Ser174 and Arg178, deleted from the final model during refinement from 5000K to 300K. (**C**) Ladder diagram of contacts within the major groove. D and A represent H-bond donors and acceptors: for adenine, A = N7 and D = N6; for thymine, A = O4; for guanine, A = N7 and O6; and for cytosine, D = N4. Bond distances given are for Br18. Lower case letters signify that the atoms are >3.5 Å away from either water or protein atoms.

**Table III.** Hin binding to mutant *hix* sites

| Base[a] | | Relative binding affinity *in vitro*[b] | | Hin dimer dissociation half-life[c] | Hin binding *in vivo*[d] | Frequency of occurrence[e] |
|---|---|---|---|---|---|---|
| | | Hin dimer | Hin-DBD | | | |
| 9 | A | 0.28 | 0.67 | ⩽3 | 0.4 | 3 |
| | **G (wt)** | **1.00** | **1.00** | **43** | **0.4** | **13** |
| | **T** | **0.14** | **0.29** | **<1** | **x** | **0** |
| | C | <0.02 | 0.026 | nd | $10^{-3}$ | 0 |
| 10 | **A (wt)** | **1.00** | **1.00** | **43** | **0.4** | **14** |
| | **G** | **0.50** | **0.36** | **16** | **x** | **2** |
| | T | <0.013 | 0.014 | nd | $10^{-6}$ | 0 |
| | C | <0.013 | 0.029 | nd | $10^{-6}$ | 0 |
| 11 | A | 0.15 | 0.14 | ⩽3 | 0.4 | 2 |
| | **G** | **1.85** | **2.50** | **125** | **0.4** | **8** |
| | **T (wt)** | **1.00** | **1.00** | **43** | **0.4** | **6** |
| | **C** | **0.039** | **0.041** | **nd** | **$10^{-4}$** | **0** |

[a]Wild-type sequence is marked by (wt); mutants for which X-ray structures are available are in bold.
[b]Ratio $K_d$(wild type)/$K_d$ for binding of the full-length Hin protein or Hin-DBD to mutant sequences. $K_d$(wild type) is 3.8 ± 0.8 and 34 ± 9 nM for full-length Hin and Hin-DBD, respectively.
[c]Time in minutes required for 50% of preformed complexes to dissociate after addition of 500-fold molar excess of wild-type *hix* sites. Values are derived for the time period after the initial rapid decay, except for G9T, where no complexes remained after 2 min (see Figure 3C). nd indicates not determined because of the very poor binding to these mutant sites.
[d]*In vivo* binding of Hin to mutant *hix* sequences based on an *in vivo* repression assay using the P22 challenge phage system. The number of lysogens per infected cell when *hin* was expressed under *tacP* control in the presence of 0.1 mM IPTG is given. Data are from Hughes *et al*. (1992). x indicates that an accurate determination could not be made for the particular mutant construct.
[e]Number of occurrences of a base at each position within half-sites of eight different DNA invertase recombination sites: *hixL*, *hixR*, *gixL*, *gixR*, *cixL*, *cixR*, *pixL* and *pixR*.

native structures Br18, I5 and I4-2, with average H-bond distances from Ser174-Oγ to A10-N7 and A10-N6 of 2.74 and 3.11 Å, respectively.

W2 is recruited to read the corresponding atoms of residue A21, 1 bp downstream. Water W2 has four protein–DNA interactions: it donates an H-bond to A21-N7 and Ser174-O, and accepts an H-bond from A21-N6 and W1. Atoms W2, A21-N7, A21-N6 and Ser174-O are essentially co-planar, whereas the W1–W2 vector is perpendicular to this plane (Figure 2A). The significance of this geometry will become apparent in discussion of the mutant T11G. The complete coordination of W2 results in its being associated stably within the Hin–DNA interface; the *B*-values of W2 are 40–45% lower than the average *B*-values in both the Br18 and I5 wild-type structures. W2 provides the only means of recognition of the T11:A21 base pair.

*Arg178 and W1*. At the other end of the recognition helix, G9-N7 is contacted directly by an H-bond donated by Arg178-Nη1. Arg178-Nη2 stabilizes the DNA–protein complex by donating an H-bond to the T8-O2P backbone atom. W1 is recruited to read the acceptor T22-O4 1 bp downstream. This water also has four interactions: three are H-bonds (with Arg178-Nε, T22-O4 and W2) and one is a van der Waals contact with T22-methyl (W1–C5A distance = 3.6 Å). The importance of this favorable interaction, which could also involve weak CH–O H-bonding (Mandel-Gutfreund *et al*., 1998), will become apparent in consideration of the mutant A10G. W1 also lacks tetrahedral geometry, as atoms W1, W2, Arg178-Nε and T22-O4 are co-planar (Figure 2A). The *B*-values for W1 are 35% (Br18) and 52% (I5) higher than for W2, indicating that W1 displays greater mobility than W2. As discussed below, the properties of the mutant binding sites and the structures of their complexes suggest that W1 is less important in the recognition process than W2.

### *Binding properties of Hin to wild-type and mutant DNA sites*

An earlier mutational analysis of each position within the *hix* recombination site identified the base pairs G9:C23, A10:T22 and T11:A21 as being important for Hin binding (Hughes *et al*., 1992). In the Hughes *et al.* study, the effects of mutations on Hin binding were evaluated by an *in vivo* repression assay (included in Table III). We have extended this analysis by comparing the Hin-binding properties of individual base pair substitutions at each of these three positions *in vitro*. Binding was evaluated both for the full-length Hin protein and for the isolated 52 amino acid DNA-binding domain used in the X-ray analysis. An intact 26 bp *hix* site, with each half-site consisting of the sequence used for crystal growth, was used for the binding measurements. Representative binding isotherms for all base pair combinations at position T11:A21 are shown in Figure 3A (full-length Hin) and B (Hin-DBD), and the binding affinities for the complete data set are given in Table III. The Hin-DBD binds non-cooperatively to each half-site with an apparent dissociation constant for the initially bound protomer of ~35 nM under the present conditions (see also Bruist *et al*., 1987; Sluka *et al*., 1990). Native Hin binds cooperatively as a dimer with a dissociation constant of ~4 nM (see also Glasgow *et al*., 1989). When preformed complexes were challenged with a 500-fold molar excess of wild-type binding sites, a small number of complexes (~15%) dissociated rapidly, but the remainder dissociated with a $T_{1/2}$ of 43 min (Figure 3C).

The mutations that displayed the most severe effect on Hin binding *in vitro*, G9C, A10C, A10T and T11C, had also been identified in the *in vivo* analysis (Table III). With the exception of T11G, all of the remaining mutants displayed reduced binding affinities and markedly faster dissociation rates (Table III; Figure 3). The binding

properties of four mutants are discussed below along with the atomic structures of the respective Hin-DBD complexes.

### *X-ray structures of mutant DNA complexes*

*Mutant T11C*. In the wild-type structure, base pair T11:A21 is only recognized via an intermediate water (W2) that both donates and accepts a H-bond to base A21. However, a T:A to C:G transition mutation at this position produces a catastrophic decrease in DNA binding, reducing binding of both the Hin dimer and DBD by a factor of ~25 (Table III). The 2.3 Å crystal structure of Hin-DBD bound to T11C reveals small changes to the DNA structure, but an absence of both W1 and W2 (Figure 4A). Ser174-Oγ remains H-bonded to A10-N7, and the Arg178 side chain is shifted slightly such that it can now make a weak H-bond with G9-O6 (3.4 Å), in addition to the normally observed bonds to G9-N7 (2.6 Å) and T8-O2P. A slight increase in propeller twisting is induced into the DNA over the A10:T23 to A12:T21 region, probably because of the proximity of T20-O4 to C11-N4, and G21-O6 to A10-N6, both lying within H-bonding distances of 3.0 and 3.1 Å, respectively.

The most striking difference from the wild-type structures is the lack of the two bridging water molecules. The resolution of T11C is similar to that of wild-type I5, although it has a slightly higher average *B*-value (60 versus 52 Å$^2$). Nevertheless, given the ability to observe W2 in the I5, T11G and A10G data sets, W2 should have been visible in the refined T11C structure, if present. Because W1 has a higher *B*-value than W2 in the wild-type structures, one cannot be absolutely certain that W1 is absent.

Why should a C:G for T:A substitution at position 11 result in destabilization of W2? Even though the DNA bases in this region are more propeller twisted, the slight change in van der Waals surfaces in the T11C complex should not interfere with the presence of a water at W2. However, such a water would be located in close proximity to three electronegative atoms. G21-O6 and G21-O7 are 3.1 Å apart, while Ser174-O is 5.4 Å from G21-O6 and 6.0 Å from G21-N7. Hence, an unfavorable electronegative chemical environment would exist for an intervening water molecule with its electronegative oxygen atom.

Loss of an ordered water molecule at W2 eliminates any interactions between Hin and base pair 11:21. A further consequence of its absence is that W1 is destabilized, since W1 is normally H-bonded to W2. Without an ordered W1, T22 is no longer read by Hin, and the Arg178 side chain shifts slightly towards base G9 to allow a weak interaction (3.4 Å) with its O6 atom. In sum, all three water-mediated contacts are lost in the T11C mutant. Loss of three of the five direct and indirect H-bonds that are normally present

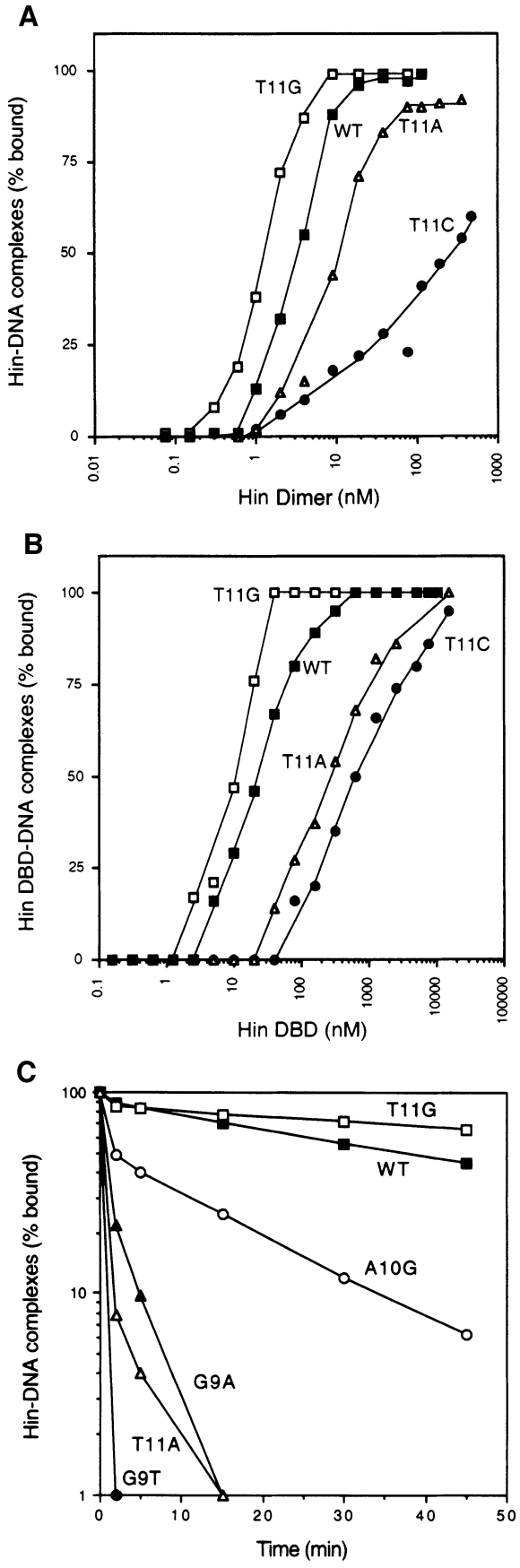
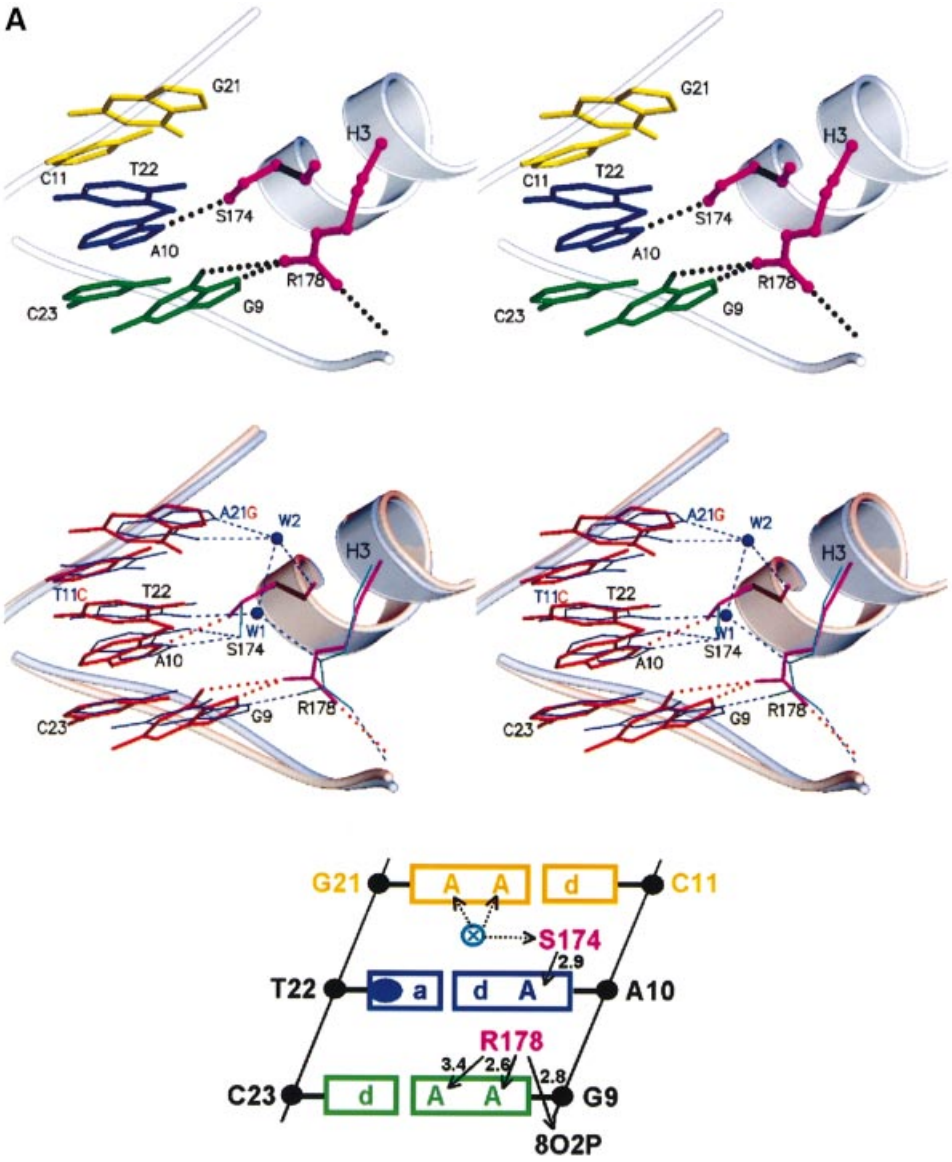



**Fig. 3.** Hin binding to mutant *hix* sites. (**A**) Binding isotherms of full-length Hin to mutant *hix* sites containing changes at position 11. (**B**) Binding isotherms of the Hin-DBD to mutant *hix* sites containing changes at position 11. Data for (A) and (B) were derived from gel mobility shift assays. (**C**) Dissociation kinetics of complexes containing Hin bound to mutant *hix* sites, measured as described in the text.

within the major groove recognition region is consistent with the large reduction in binding affinity.

*Mutant T11G*. T11G binds Hin slightly better than the wild-type site, as shown by the increased relative binding affinity for both the DBD and intact Hin, and by the longer dissociation half-life (Table III; Figure 3). Consistent with efficient binding, a G:C at this position is the most frequently represented base pair among the half-sites of four well characterized inversion systems (Table III), and is present at the other half-site of wild-type *hixL*. The X-ray structure of T11G shows only a minor perturbation when compared with the wild type (Figure 4B). The substituted base pair G11:C21 shifts slightly along its long axis toward strand 1 of the DNA helix in order to improve C-upon-T and G-upon-A stacking with A10:T22, but the helix structure is otherwise nearly unaltered. Base pair angle parameters and positions of H-bond donors and acceptors outside of the substituted base pair are virtually identical between structures T11G and Br18. Even though the resolution of the T11G structure is low (2.83 Å), weak electron density at the position of W2 is visible in an $F_o - F_c$ map contoured at 3.5σ. The *B*-value for W2 in T11G is ~40% lower than the average *B*-value for the structure, similar to W2 in the wild-type structures. The H-bond between W2 and A21-N7 in the wild type is lost and cannot be replaced with one to G11-O6, which is 5.5 Å away from W2. Thus, the H-bonds between C21-N4 and W2, and between W2 and Ser174-O, appear sufficient to preserve the binding of water W2.

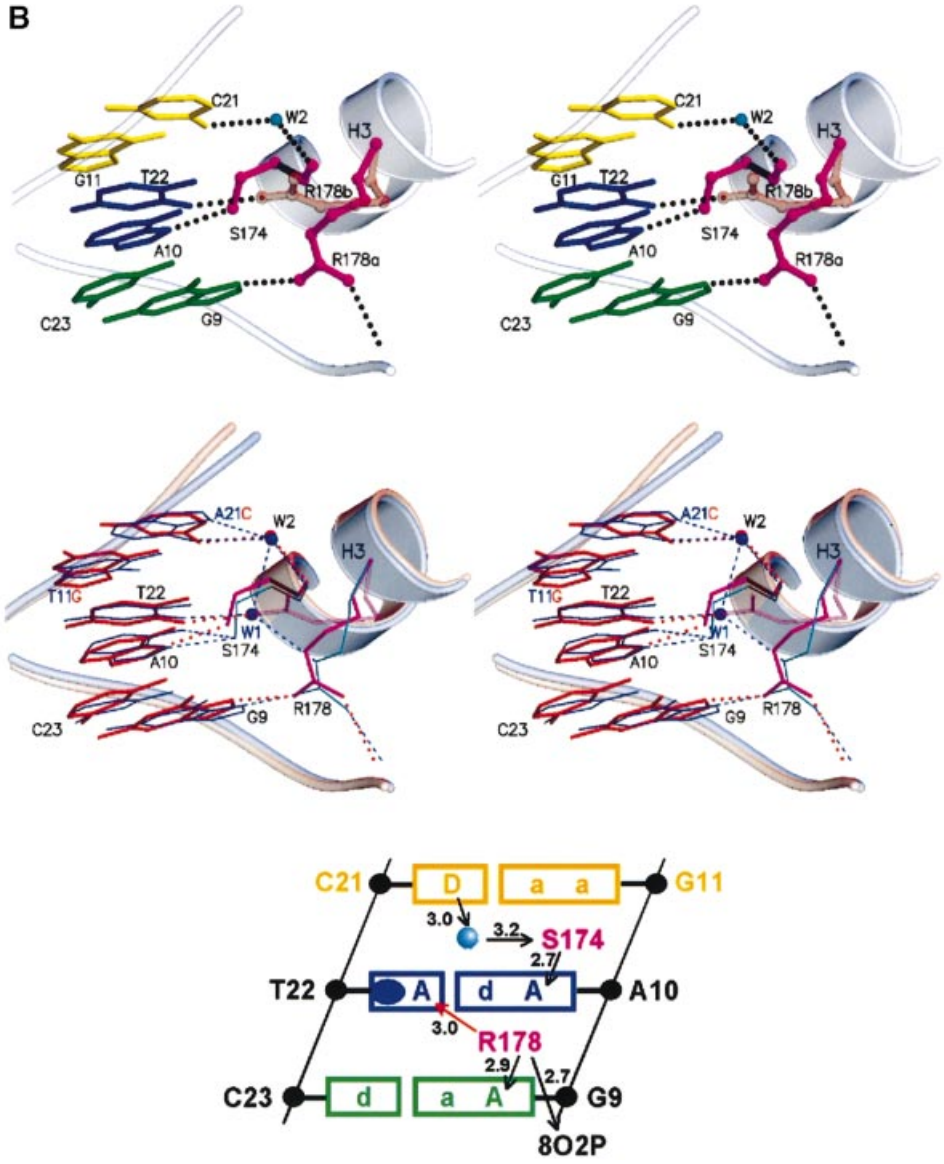
Water W1 is not observed in this mutant. Instead, the electron density map unexpectedly shows that the side chain of Arg178 adopts two distinct conformations. One conformation maintains wild-type interactions with G9-N7 and T8-O2P, whereas in the other less well defined conformation, the arginine side chain swings upward into the position occupied by W1 in the wild-type structures, and donates an H-bond directly to T22-O4 (2.9 Å) (Figure 4B). In this second conformation, the arginine guanidinium group is co-planar with base T22 and makes



van der Waals contact with the thymine methyl, but is both too far away and in the wrong orientation to donate an H-bond to water W2 (Nη1 to W2 = 3.8 Å). The second conformation of Arg178 is less well ordered since both its real-space *R*-value (0.43) and average *B*-value (50 Å$^2$) are much higher than for the wild-type conformation (0.19 and 25 Å$^2$, respectively). We would expect that W1 would be present when Arg178 is in its wild-type conformation, but the resolution of the data is insufficient for this to be determined.

*Mutant A10G*. Binding affinities measured for A10G are only 2- to 3-fold lower for the Hin dimer and DBD, but the complex appears significantly less stable (Table III; Figure 3). The X-ray structure of the complex at 2.24 Å shows very little alteration to the local structure of DNA (Figure 4C). Base pairs G10:C22 and T11:A21, and water W2, all shift slightly toward strand 1 of the helix, but the rest of the DNA structure is largely unchanged. Direct and water-mediated interactions through W2 between Ser174 and DNA are nearly identical to wild type. Whereas clear density is present for W2, there is no evidence for an ordered water at the W1 position. We would have expected to be able to observe density for W1 given the resolution of the data, albeit the average *B*-value for atoms in the A10G structure is higher than the wild-type structures. However, the presence of negative density in an ($F_{o,Br\ or\ I5} - F_{o,A10G}$)$e^{\phi Br\ or\ I5}$ map contoured at 3.5σ at the position of W1 suggests that W1 is indeed absent in A10G. Consistent with the absence of W1, the Arg178 side chain is shifted slightly towards base C22 but remains H-bonded to G9.

Why is an ordered water absent from the W1 position in A10G? The X-ray structure shows that C22-N4, W2 and Arg178-Nε could in principle H-bond with a water molecule at W1. The direction of the H-bond between W1 and the new C22-N4 atom would be reversed from the wild-type situation with T22-O4, but this in itself should not be damaging. However, the methyl group of base T22 that is in van der Waals contact with W1 in the wild-type structure is absent from A10G. This methyl helps to bury



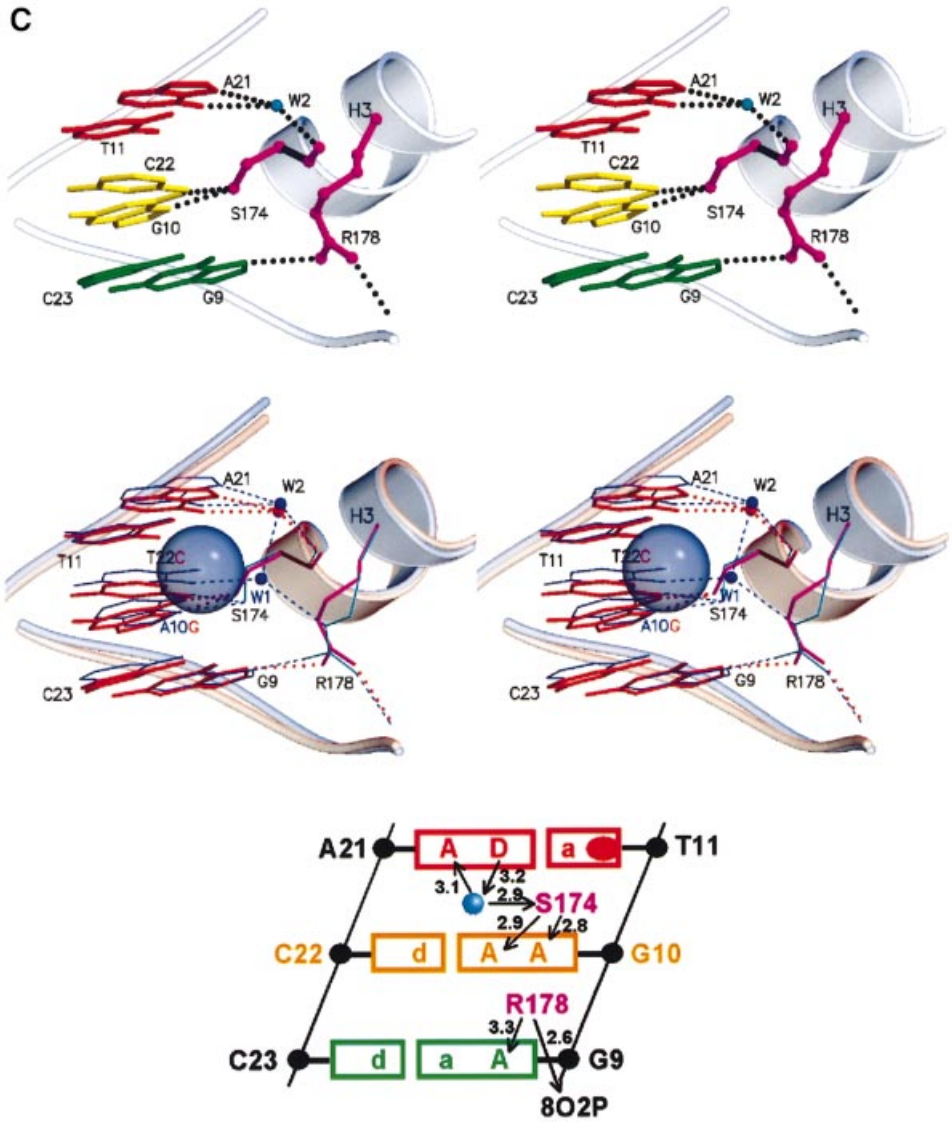
W1 and thus slows its exchange with bulk solvent, enabling it to participate in sequence recognition. We suggest that loss of this thymine methyl group destabilizes W1, which results in a modest reduction in binding.

*Mutant G9T*. G:C to T:A substitution at base pair 9 reduces the affinity of the Hin-DBD and dimer 3.5- or 7-fold, respectively (Table III). The complexes formed with the Hin dimer bound to G9T are extremely unstable as they immediately exchange upon addition of competitor (Figure 3). The X-ray structure again reveals little change in local structure of DNA except for a slight rotation of the adenine of the T9:A23 base pair (Figure 4D). Superimposition of the wild-type structure shows that the H-bond donor and acceptors on the DNA are positioned almost identically except for the mutant base pair. Neither W1 nor W2 is visible in the 2.86 Å electron density maps, but this can probably be ascribed to the low resolution. W2 would be expected, however, since there is no apparent change in either protein or DNA structure around its site. As observed in the other structures, Ser174-Oγ contacts A10 directly. The major new feature of the G9T mutant structure is addition of a T9-methyl. Whereas the T9-O4 position superimposes almost perfectly on the G9-O6 of wild type, the thymine methyl blocks interaction of this atom with Arg178, and repositions the arginine side chain away from the base. van der Waals interaction between Arg178 and the thymine methyl may partially compensate for the loss of a direct H-bond with the base.

Because of the repositioning of the Arg178 side chain by the mutant T9 methyl group, the Nε of the Arg178 side chain is no longer able to H-bond to W1. Thus, a water at the position of W1 would have only three interactions: H-bonds with T22-O4 and W2 (assuming it were present), and van der Waals contact with T22-methyl. As was observed with A10G, the loss of one interaction would be predicted to destabilize W1. The loss of W1 would be expected to reduce even further the efficacy of Hin binding to G9T.

## Discussion

The properties of mutant *hix* sites indicate that the native sequence is exquisitely optimized for Hin binding. Only



one of the base pair substitutions within the G9–A10–T11 recognition region, T11G, produces as effective a substrate as the wild-type sequence. The crystal structures of four mutant complexes have given us a detailed picture of selective recognition by Hin within the major groove. In particular, this study has provided direct structural support for the importance of specific water molecules, and highlighted the importance of thymine methyls in both stabilizing intermediate waters and in positioning amino acid side chains.

### *Importance of water and thymine methyls for Hin–DNA recognition within the major groove*

The role of intermediate water is most dramatically illustrated at base pair 11:21. Recognition at this position is achieved solely through an intermediate water molecule, yet base pair changes at this position can have enormous effects on Hin binding. The T11C structure reveals that a very small change in the chemical environment can destabilize an ordered water, resulting in a drastic decrease in binding affinity. In T11C, substitution of the A21-N6

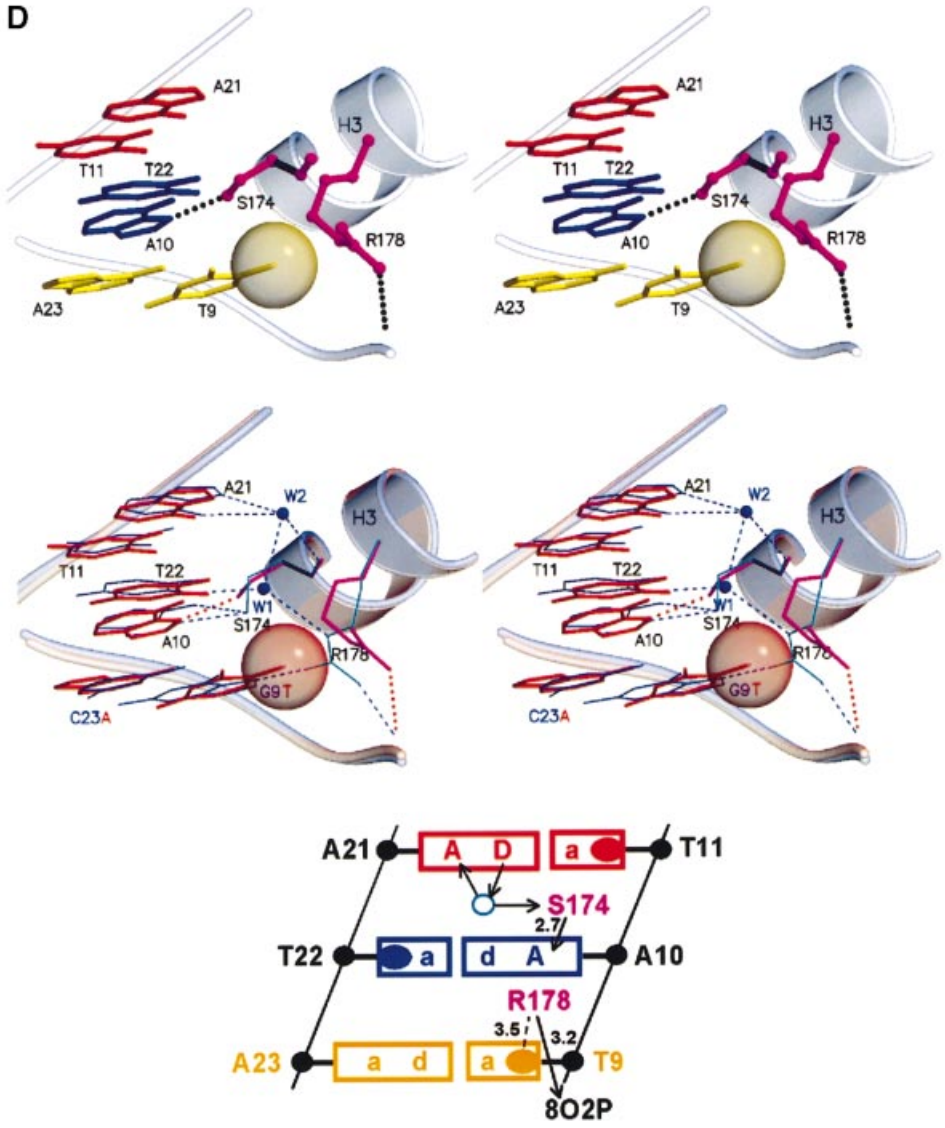



**Fig. 4.** X-ray structure of Hin-DBD bound to mutant *hix* sites. For each part, the top panel is a stereo view of the major groove from G9:C23 to C11:G21 as in Figure 2, but the mutant base pair is colored gold. The middle panel is a stereo view of an r.m.s. fitting between the mutant (red) and wild-type Br18 (blue) structures. The bottom panel is a ladder diagram of the major groove contacts observed in the mutant structure. (**A**) T11C–Hin-DBD complex (2.28 Å). The circled X and three dotted arrows in the ladder diagram mark the empty site of W2, which is destabilized because it would be located between three electronegative H-bond acceptors. Loss of W2, in turn, destabilizes binding of W1, so that this water is also unobserved. (**B**) T11G–Hin-DBD complex (2.83 Å). The Arg178 side chain has two conformations, one similar to wild type and the other swings up to contact T22-O4, thereby excluding W1. (**C**) A10G–Hin-DBD complex (2.24 Å). The wild-type T22-methyl, which is not present in the A10G mutant, is rendered as a van der Waals sphere in the superimposition. The absence of W1 in the A10G complex probably reflects the lack of the van der Waals contact with a thymine methyl from base 22. (**D**) G9T–Hin-DBD complex (2.86 Å). The methyl group of mutant base T9 is rendered as a van der Waals sphere in the structure panels. An open circle in the ladder diagram signifies that W2 is expected but is not observed, probably because of the relatively low resolution of the X-ray data. The dashed line denotes a van der Waals contact between Arg178 and T9-methyl. This methyl group of mutant base T9 repositions the Arg178 side chain such that it can no longer H-bond with base 9 or with W1.

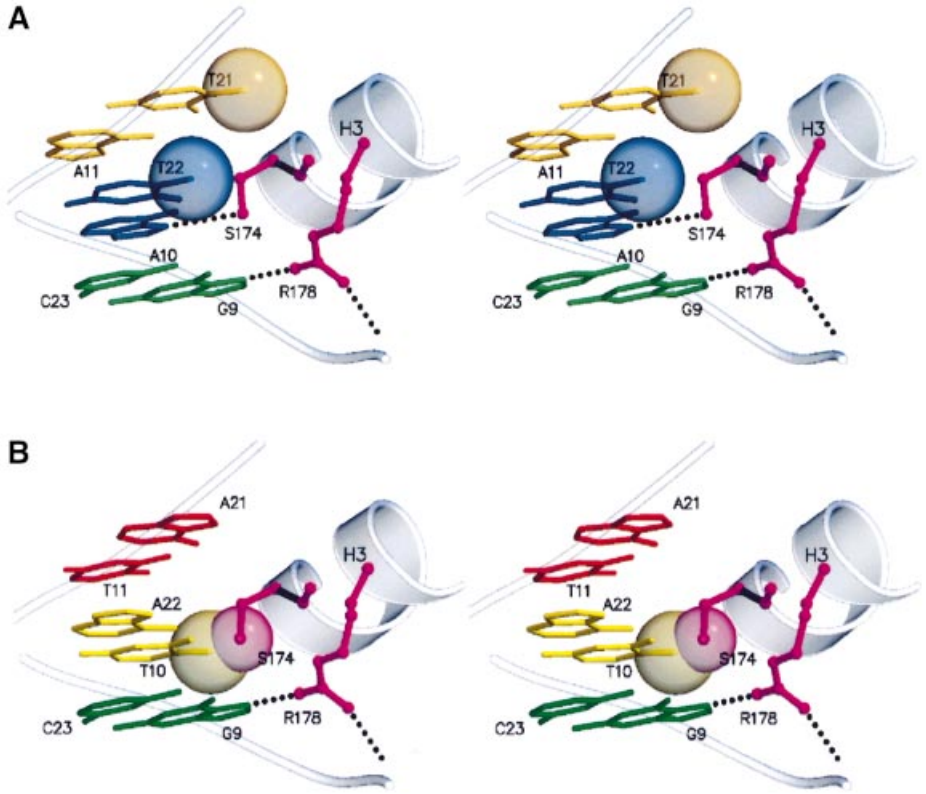



**Fig. 5.** Hin–DNA interactions in models of T11A and A10T. (**A**) Energy-minimized model of T11A. The T21 and T22 methyl groups are rendered as van der Waals spheres. The T21-methyl occludes W2 but may stabilize W1. (**B**) Substitution of a T:A base pair at position 10 in mutant A10T results in a clash with the Ser174 side chain. The van der Waals surfaces of T10–C5A and Ser174-Oγ are rendered. Substitution of a C:G base pair in mutant A10C also results in a clash with the Ser174 side chain (not shown). Note that saturable binding was never achieved with A10T or A10C, even at native Hin or Hin-DBD concentrations exceeding 1 μM. Models were generated by visually fitting the substituted base pair into the Br18 structure, and for T11A, energy minimized using CNS (Brünger *et al*., 1998).

atom by G-O6 leads to the loss of W2, which, in turn, destabilizes W1. When an A:T base pair is substituted at position 11, W2 is also predicted to be lost, as shown in the energy-minimized model of mutant T11A in Figure 5A. In this case, a water at the W2 position could be located between the two H-bond acceptors T22-O4 and Ser174-O. However, this water would be 3 Å away from the T21-methyl, which would effectively exclude it from this location. Whereas the absence of W2 in T11A might be expected to destabilize W1, as observed in T11C, the mobility of W1 may actually be restrained by the T21-methyl 4 Å away (illustrated in Figure 5A). Thus, the less severe binding defect by T11A (Table III; Figure 3), as compared with T11C, is predicted to arise from loss of only water W2. The structure of T11G (Figure 4B) reinforces the importance of W2 and illustrates that altered H-bonding profiles can be accommodated. T11G, which is a high-affinity substrate for Hin binding, has a well ordered W2 water that is H-bonded to the mutant C21-N4 atom and Ser174-O.

Relative *B*-values suggest that W1 is less well ordered than W2, and the combination of the structural and binding properties of A10G implies that W1 plays a less important role in binding than W2. The structure of the A10G complex is essentially identical to the wild type except for the absence of W1 and the methyl at base 22. However, binding affinities measured for A10G are reduced 2- to 3-fold, and the complexes are markedly less stable. To our knowledge, the A10G mutant structure is the first example where a specific thymine methyl has been shown to stabilize an intermediate water. A somewhat analogous situation at the protein level occurs within the recognition helix of Engrailed and other homeodomains where the methyl of a conserved alanine stabilizes a set of waters within the DNA interface (Fraenkel *et al*., 1998).

The thymine methyl introduced in mutant G9T has also proved to be the dominant structural feature affecting Hin–DNA interaction. The van der Waals surface of the methyl blocks the ability of Arg178 to H-bond to the T9-O4 and redirects the side chain away from the major groove. Consequently, the Arg178-Nε is not in a position to H-bond with W1.

### *Surface complementarity and flexibility of side chains within the protein–DNA interface*

A final point comes from considering why a pyrimidine at position 10 is so detrimental for Hin binding. Indeed, C:G or T:A base pair substitutions nearly abolish Hin binding, and a pyrimidine is never observed at this position at any of the DNA invertase recombination sites (Table III). Initial model building of complexes with C:G or T:A substitutions at position 10 revealed a severe clash between Ser174-Oγ and C10-C5, and even worse, between Ser174-Oγ and T10-C5A (Figure 5B). The rigidity of the

Ser174 side chain would require large changes in the DNA and/or protein structure, as shown by energy-minimization calculations. The effect of the rigid Ser174 side chain contrasts with the flexibility of the Arg178 side chain. Arg178 in the T11G structure was found to exist in a second conformation in which its guanidinium group was H-bonded to T22 instead of G9. Likewise, the repositioning of the Arg178 side chain by the introduced thymine methyl in the G9T structure occurred without global changes to the structure.

## Materials and methods

### *Crystallization*

Oligonucleotides were synthesized by solid-phase phosphoramidite chemistry, purified on 20% polyacrylamide gels, and annealed at 2 mM in the presence of 100 mM NaCl and 20 mM sodium cacodylate pH 7.0. The synthetic 52 amino acid C-terminal Hin-DBD peptide used in the previous study (Feng *et al*., 1993, 1994) was employed for all crystal growth in this work except I7/19, in which the peptide used was purchased from Genemed Inc. (San Francisco, CA). CD analysis showed that the new peptide is unfolded in the absence of DNA. Crystals were grown by sitting drop vapor diffusion at 4°C (Form 1) and 21°C (Form 2) by combining pre-incubated DNA–Hin-DBD complex with reservoir solution. The DNA/Hin mole ratios were 1:2.5, 1.7:1 and 2:1 for Form 1 derivative, Form 1 native and mutant, and Form 2 crystals, respectively. The corresponding volume ratios of complex and reservoir solution were 1:2, 1:5 to 8 and 1:3, respectively. The Form 1 reservoir solution had 100 mM NaCl/$CaCl_2$/Tris pH 8.5 and 25% PEG400. Form 2 reservoir solution had 20–100 mM $MgCl_2$, 100 mM NaOAc pH 4.6 and 25% PEG400.

### *Data collection, phasing and refinement*

The Hin–DNA co-crystals all belonged to space group $C222_1$, with Form 1 unit cell parameters ~86 × 82 × 45 Å and Form 2 cell parameters ~66 × 69 × 62 Å. Both crystal forms contained one complex per crystallographic asymmetric unit. Data at 100°K were collected at UCLA (I4, I5, Nat-2 and I4-2), at Chess beamline A1 (Nat-1 and T11C), and at NSLS beamlines X25 (G9T, A10G and T11G) and X12C (I7/19, I4, I5 and Br18), and were processed with HKL (Otwinowski, 1993). Phasing, density modification and model building were carried out with SOLVE 1.16 (Terwilliger and Berendzen, 1999), DM (Cowan, 1994) and O (Jones *et al*., 1991), respectively. The structures were refined using CNS 1.0 (Brünger *et al*., 1998), applying both bulk solvent and overall anisotropic temperature factor corrections. The force fields of Engh and Huber (1991) were used for protein and those of Parkinson *et al*. (1996) for DNA.

The least confident parts of the earlier model, bases at each end of the DNA duplex and the C-terminal tail of the Hin peptide, were now observed to adopt a different conformation in the new heavy atom phased experimental maps. In the earlier model, the C-terminal residues had the lowest density correlation and density index as determined with SFCHECK (Vaguine *et al*., 1999), and in hindsight the non-continuous density was incorrectly traced as a peptide chain. To minimize bias in the current models, as well as to further support the new tracing, these residues were omitted in the initial models. Upon refinement using CNS 1.0, $2F_o - F_c$ and $F_o - F_c$ maps showed clear density consistent with the new tracing. The new conformation of the C-terminal tail observed in Form 1 crystals was confirmed with SIRAS phases for Form 2 crystals. The final models consist of DNA residues 2–29, protein residues 139–185 or 187, and a small number of solvent molecules. All of the protein residues are within the most favored or additionally allowed conformations as calculated by PROCHECK (Laskowski *et al*., 1993). Statistics for each of the structures are listed in Tables I and II, along with PDB accession codes for the deposited coordinates and structure factors. Complete details of the X-ray structures can be found in Chiu (2001). Molecular figures were rendered using the Insight II software package (Accelrys, Inc.), Ribbons 3.0 (Carson, 1987) or with Molscript 1.4 (Kraulis, 1991), Raster3D (Merritt and Bacon, 1997) and ImageMagick©.

### *DNA-binding assays*

Wild-type and mutant Hin-binding sites were cloned between the *Eco*RI and *Sal*I sites of pBR322 as intact *hix* sites. In the wild-type sequence, TCTT**ATC**AAAAACAATTGTTTTT**GAT**AAGA, the underlined sequence denotes the 14 bp segment present in the crystal structures and the bold sequences are the 3 bp segments that contain substitutions in the mutants. $^{32}$P-end-labeled probes used for the DNA-binding assays were prepared by PCR and contained ~50 bp of flanking pBR322 DNA on each side of the *hix* site. Full-length Hin was purified as described in Merickel *et al*. (1998) and the DBD used for the binding reactions was from Genemed Inc. Binding reactions were performed in 20 µl of 20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM DTT, 5 mM $MgCl_2$, 10% v/v glycerol, 25 µg/ml poly(dI):(dC) and CHAPS at 10 mM for the Hin-DBD and 20 mM for the native Hin. After a 20 min incubation at 23°C with different concentrations of Hin or Hin-DBD, the binding reaction was loaded onto a polyacrylamide gel (6% 29:1 acrylamide:bisacrylamide for Hin and 8% 19:1 acrylamide:bisacrylamide for Hin-DBD) and subjected to electrophoresis at 15 mA. The free DNA and Hin-bound DNA complexes were quantified by phosphoimaging, and $K_d$ values were determined by plotting log ($b$/1 – $b$) against log [P], where $b$ is the fraction of labeled probe bound and P is protein concentration, and the $x$-intercept was determined. $K_d$ was then calculated according to the equation $K_d = 10^{x\text{-intercept}}$. Values in Table III were averaged from at least five independent measurements for each mutant relative to wild type and standard deviations were typically ⩽25%.

Dissociation kinetics were measured by preforming Hin–DNA complexes in scaled-up reactions using sufficient Hin to assemble 50–75% of the labeled probe into complexes. Annealed 34 base oligonucleotides representing the wild-type *hix* site were added at ~500-fold molar excess over the probe. Samples were loaded onto an electrophoresing polyacrylamide gel immediately before and at various times after addition of competitor.

## Acknowledgements

We thank Dr Duilio Cascio for help with synchrotron data collection on the native and mutant DNA crystals, Dr Cascio and Dr Kenneth Goodwill for collecting synchrotron datasets I4, I5 and Br18, and Dr Michael Sawaya for collecting synchrotron dataset I7/19. We also thank Dr Cascio for many helpful discussions during the early part of this project, Dr Rick Fahrner for help with SOLVE, Dr Sawaya for help with general crystallographic questions and Dr Gerson Cohen for help with Ribbons. Synthetic Hin peptide used for most of the crystal growth was generously provided by S.Horvath, P.Dervan and M.Simon (Caltech). This research was carried out with the support of the National Institute of Health, and a Dissertation Year Fellowship to T.K.C. from the Department of Chemistry and Biochemistry at UCLA.

## References

Bruist,M.F., Horvath,S.J., Hood,L.E., Steitz,T.A. and Simon,M.I. (1987) Synthesis of a site-specific DNA-binding peptide. *Science*, **235**, 777–780.

Brünger,A.T. *et al*. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.

Carson,M. (1987) Ribbon models of macromolecules. *J. Mol. Graphics*, **5**, 103–106.

Chiu,T.K. (2001) How Hin recombinase, Fis and cations bind DNA. Chapter 4. PhD thesis, Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, CA.

Cowan,K. (1994) DM: an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newslett. Protein Crystallogr.*, **31**, 34–39.

Dickerson,R.E. and Chiu,T.K. (1997) Helix bending as a factor in protein–DNA recognition. *Biopolymers*, **44**, 361–403.

Engh,R.A. and Huber,R. (1991) Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr. A*, **47**, 392–400.

Feng,J.A., Simon,M., Mack,D.P., Dervan,P.B., Johnson,R.C. and Dickerson,R.E. (1993) Crystallization and preliminary X-ray analysis of the DNA binding domain of the Hin recombinase with its DNA binding site. *J. Mol. Biol.*, **232**, 982–986.

Feng,J.A., Johnson,R.C. and Dickerson,R.E. (1994) Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science*, **263**, 348–355.

Fraenkel,E. and Pabo,C.O. (1998) Comparison of X-ray and NMR structures for the Antennapedia homeodomain–DNA complex. *Nature Struct. Biol.*, **5**, 692–697.

Fraenkel,E., Rould,M.A., Chambers,K.A. and Pabo,C.O. (1998) Engrailed homeodomain–DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures. *J. Mol. Biol.*, **284**, 351–361.

Gehring,W.J., Qian,Y.Q., Billeter,M., Furukubo-Tokunaga,K., Schier, A.F., Resendez-Perez,D., Affolter,M., Otting,G. and Wuthrich,K. (1994) Homeodomain–DNA recognition. *Cell*, **78**, 211–223.

Glasgow,A.C., Bruist,M.F. and Simon,M.I. (1989) DNA-binding properties of the Hin recombinase. *J. Biol. Chem.*, **264**, 10072–10082.

Grant,R.A., Rould,M.A., Klemm,J.D. and Pabo,C.O. (2000) Exploring the role of glutamine 50 in the homeodomain–DNA interface: crystal structure of engrailed (Gln50 → Ala) complex at 2.0 Å. *Biochemistry*, **39**, 8187–8192.

Grindley,N.D.F. (1993) Resolvase-mediated site-specific recombination. In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Vol. 8. Springer-Verlag, Berlin, Germany, pp. 236–267.

Hirsch,J.A. and Aggarwal,A.K. (1995) Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. *EMBO J.*, **14**, 6280–6291.

Horton,N.C. and Perona,J.J. (1998) Recognition of flanking DNA sequences by *Eco*RV endonuclease involves alternative patterns of water-mediated contacts. *J. Biol. Chem.*, **273**, 21721–21729.

Hughes,K.T., Gaines,P.C., Karlinsey,J.E., Vinayak,R. and Simon,M.I. (1992) Sequence-specific interaction of the *Salmonella* Hin recombinase in both major and minor grooves of DNA. *EMBO J.*, **11**, 2695–2705.

Joachimiak,A., Haran,T.E. and Sigler,P.B. (1994) Mutagenesis supports water mediated recognition in the Trp repressor–operator system. *EMBO J.*, **13**, 367–372.

Johnson,R.C. (2002) Bacterial site-specific DNA inversion systems. In Craig,N., Craigie,R., Lambowitz,A. and Gellert,M. (eds), *Mobile DNA II*. ASM Press, Washington, DC, pp. 230–271.

Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lawson,C.L. and Carey,J. (1993) Tandem binding in crystals of a Trp repressor/operator half-site complex. *Nature*, **366**, 178–182.

Li,T., Stark,M.R., Johnson,A.D. and Wolberger,C. (1995) Crystal structure of the MAT**a**1/MATα2 homeodomain heterodimer bound to DNA. *Science*, **270**, 262–269.

Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.

Mandel-Gutfreund,Y., Margalit,H., Jernigan,R.L. and Zhurkin,V.B. (1998) A role for CH–O interactions in protein–DNA recognition. *J. Mol. Biol.*, **277**, 1129–1140.

Merickel,S.K., Haykinson,M.J. and Johnson,R.C. (1998) Communication between Hin recombinase and Fis regulatory subunits during coordinate activation of Hin-catalyzed site-specific DNA inversion. *Genes Dev.*, **12**, 2803–2816.

Merritt,E.A. and Bacon,D.J. (1997) Raster3D photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.

Nadassy,K., Wodak,S.J. and Janin,J. (1999) Structural features of protein–nucleic acid recognition sites. *Biochemistry*, **38**, 1999–2017.

Nadassy,K., Tomas-Oliveira,I., Alberts,I., Janin,J. and Wodak,S.J. (2001) Standard atomic volumes in double-stranded DNA and packing in protein–DNA interfaces. *Nucleic Acids Res.*, **29**, 3362–3376.

Olson,W.K., Gorin,A.A., Lu,X.J., Hock,L.M. and Zhurkin,V.B. (1998) DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163–11168.

Otwinowski,Z. (1993) Oscillation data reduction program. In Sawyer,L., Isaacs,N. and Bailey,S. (eds), *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*. SERC Daresbury Laboratory, Warrington, UK, pp. 56–62.

Otwinowski,Z., Schevitz,R.W., Zhang,R.G., Lawson,C.L., Joachimiak,A., Marmorstein,R.Q., Luisi,B.F. and Sigler,P.B. (1988) Crystal structure of Trp repressor/operator complex at atomic resolution. *Nature*, **335**, 321–329.

Parkinson,G., Vojtechovsky,J., Clowney,L., Brünger,A.T. and Berman, H.M. (1996) New parameters for the refinement of nucleic acid containing structures. *Acta Crystallogr. D*, **52**, 57–64.

Robinson,C.R. and Sligar,S.G. (1996) Participation of water in Hin recombinase–DNA recognition. *Protein Sci.*, **5**, 2119–2124.

Schwabe,J.W. (1997) The role of water in protein–DNA interactions. *Curr. Opin. Struct. Biol.*, **7**, 126–134.

Seeman,N.C., Rosenberg,J.M. and Rich,A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.

Silverman,M., Zieg,J., Mandel,G. and Simon,M. (1981) Analysis of the functional components of the phase variation system. *Cold Spring Harb. Symp. Quant. Biol.*, **45**, 17–26.

Sluka,J.P., Horvath,S.J., Glasgow,A.C., Simon,M.I. and Dervan,P.B. (1990) Importance of minor-groove contacts for recognition of DNA by the binding domain of Hin recombinase. *Biochemistry*, **29**, 6551–6561.

Terwilliger,T.C. and Berendzen,J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D*, **55**, 849–861.

Tucker-Kellogg,L., Rould,M.A., Chambers,K.A., Ades,S.E., Sauer,R.T. and Pabo,C.O. (1997) Engrailed (Gln50→Lys) homeodomain–DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity. *Structure*, **5**, 1047–1054.

Vaguine,A.A., Richelle,J. and Wodak,S.J. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D*, **55**, 191–205.

vonHippel,P.H. and Berg,O.G. (1986) On the specificity of DNA–protein interactions. *Proc. Natl Acad. Sci. USA*, **83**, 1608–1612.

Wilson,D.S., Guenther,B., Desplan,C. and Kuriyan,J. (1995) High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell*, **82**, 709–719.

Woda,J., Schneider,B., Patel,K., Mistry,K. and Berman,H.M. (1998) An analysis of the relationship between hydration and protein–DNA interactions. *Biophys. J.*, **75**, 2170–2177.

Wolberger,C. (1996) Homeodomain interactions. *Curr. Opin. Struct. Biol.*, **6**, 62–68.

Yang,W. and Steitz,T.A. (1995) Crystal structure of the site-specific recombinase γδ resolvase complexed with a 34 bp cleavage site. *Cell*, **82**, 193–207.

*Received October 12, 2001; revised and accepted December 13, 2001*